

“Mathematical Contributions to the Theory of Evolution. XII.
—On a Generalised Theory of Alternative Inheritance, with
Special Reference to Mendel’s Laws.” By KARL PEARSON,
F.R.S. Received September 11,—Read November 26, 1903.

(Abstract.)

(1). A great deal of interest has been recently excited by Mendelian theories of inheritance, and a considerable amount of controversy has arisen with regard to their relation to the statistical treatment of inheritance within large populations. The fundamental conception of the Mendelian theory is the non-fusal of the gametes of two pure races on crossing. If A, A' represent the pure gametes of one race and a, a' those of the other, the individual arising from the zygote formed by crossing these races contains not fused gametes, but pure gametes of the two original races, and these gametes are equally represented on further crossings among the hybrids. The whole may be exhibited symbolically as

$$(AA') \times (aa') = (A + A')(a + a') = (Aa) + (Aa') + (A'a) + (aa'),$$

where in the first product we have the parental zygotes; in the second product the breaking-up into the gametes is indicated, and on the extreme right the zygotes of the resulting offspring are given. This is a perfectly simple conception. We can generalise it at once by replacing the simple couplet (AA') by n such couplets, and supposing the original parental zygotes to be of the form

$$(A_1A_1) + (A_2A_2) + (A_3A_3) + \dots + (A_nA_n),$$

$$(a_1a_1) + (a_2a_2) + (a_3a_3) + \dots + (a_na_n),$$

where each corresponding pair of couplets on crossing obeys the above symbolic formula. If two individuals, constituted as above, be crossed, they give a generation of hybrids. If these hybrids now breed at random and are equally fertile among themselves, segregation takes place. If the process of random mating with equal fertility be continued generation by generation, what further changes, if any, take place, and what are the laws of inheritance within such a population? This is the general problem I have set myself to attack by mathematical analysis. I speak of it as a generalised theory of alternative inheritance; it is based on a pure gamete conception. It is not Mendelian in so far as it replaces his simple couplet by n -couplets, and supposes the character determined by these n -couplets and not a single couplet. Further, it entirely dispenses with Mendel’s fundamental ideas of recessive and dominant characters. To avoid any such assumption I speak of the protogenic and allogenic couplets (aa) and (AA) . I take

the character to be fixed in some way by the number of protogenic or allogenic couplets in the constitution of the individual zygote. It would have been equally easy to fix it by the heterogenic couplets. My object is to discover the chief features of heredity in a population breeding at random under a generalised system of pure gametes, and to determine how far the results are in accordance with the biometric experience of what happens in such populations.

I am quite aware that my theory, although far more general than Mendel's, is not the most general which might be propounded. A whole crop of Mendelian formulæ have been proposed for isolated special cases by various writers. But I venture to think that this has been done with a certain want of scientific responsibility. Their authors do not seem to have realised that every such formula involves absolutely rigid laws for inheritance within a population mating at random, and that before such formulæ are propounded, a research ought to be made as to what general laws of heredity flow from them and whether such laws are in accordance with existing experience. The present memoir will indicate that to follow up one such general Mendelian theory requires a great deal of mathematical labour, and that in future the onus of testing the generality of a Mendelian formula propounded to account for some single hybridisation result ought to fall on its propounder.

What the present investigation does demonstrate is this, that the most general theory of the pure gamete hitherto developed leads to broad features of inheritance, which are in perfect accord with the biometric theory of inheritance in populations. Nor, if this theory of the pure gamete were correct, is there anything to be surprised at in the result. The biometric theory is purely a statistical description based upon experience and is not dependent upon any physiological hypothesis. On the other hand, any physiological hypothesis applied to heredity must fall if it does not accord with the results obtained by observations on such populations.

(2). Developing the theory described above we find :—

(a). That segregation occurs when the hybrids cross.

(b). That when the members of this segregating generation cross at random the population accurately reproduces itself, and supposing no artificial, natural or reproductive selection to take place, a stable population or "race" is created, which is permanent and shows a permanent proportional frequency for each sub-class of the population. There is no room for further segregation or for "mutations" as long as this state of affairs is maintained. Any selection, however, would at once produce a progressive change in the population.

(c). The regression of the offspring on any ancestor whatever is linear.

(d). The ancestral correlations form a geometrical series. The

results (a) to (d) involve what I have termed the law of ancestral heredity, which is thus seen to hold for a general population obeying the most general theory of the pure gamete hitherto developed mathematically.

(e). The distribution of the population itself follows a skew binomial law, which approaches closer and closer to the Gaussian distribution as we increase the number of couplets in the constitution of the zygote. All these results are in broad agreement with what biometric investigations have shown us to hold for the inheritance and distribution of frequency in populations.

(3). We now turn to points of divergence :—

(f). The parental correlation is $\frac{1}{3}$.

(g). The ratio of diminution of ancestral correlation at each stage is $\frac{1}{2}$.

(h). The correlation between brothers is sensibly 0·4.

One-third is the value originally given by Francis Galton for parental correlation ; $\frac{1}{2}$ is the value of the ratio of ancestral diminution which really follows from his geometrical series $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$, etc., for ancestral contributions, and 0·4 is the corresponding value of fraternal correlation.*

The generalised Mendelian theory thus fixes absolutely inheritance for all characters in all races, *i.e.*, inheritance is no function of the number of couplets determining the constitution. These results, while in accordance with Galton's original views, do not seem in accordance with more recent statistical observations on populations.

The value for parental correlation is not constant either for character or species ; it varies sensibly, and it clusters about a value 0·45 to 0·5, sensibly higher than the value $\frac{1}{3}$ above given. The value for fraternal correlation is, indeed, somewhat larger than that for parental correlation, but it clusters about a value 0·5 to 0·6, and is sensibly higher than the value 0·4 obtained from this Mendelian theory. The values $\frac{1}{6}$, $\frac{1}{12}$, $\frac{1}{24}$, found for grandparental, great grandparental and great great grandparental correlations from this theory are all very sensibly less than the values actually obtained from observation. The ratio of diminution is more nearly $\frac{2}{3}$ than $\frac{1}{2}$.

Hence, this Mendelian generalisation fails when we test it by actual numbers.

(4). Points of theoretical divergence, not depending upon numerical measures, are the following :

(i). If p be the number of allogenic couplets in the father, q in the mother, the number m_{pq} to be expected on the average in the offspring is given by

$$m_{pq} = \frac{1}{9} \frac{(n+2p)(n+2q)}{n} \dots\dots\dots (i)$$

* 0·4 and not $\frac{2}{3}$ as in "Natural Inheritance" is the value for fraternal correlation which flows from Galton's hypothesis, see 'Roy. Soc. Proc.,' vol. 62, p. 410.

on the generalised pure gamete theory ; and by

$$m_{pq} = \frac{1}{12}n + \frac{1}{2}(p + q) \dots\dots\dots (ii)$$

on the theory of linear regression.

In other words, regression holds for the great bulk of a Mendelian population, but grows gradually less as we treat the matings of nearly pure allogenic parents, ceasing entirely with absolutely pure allogenic parents. When $n = 5$, a mating between pure allogenic parents occurs only once in a million matings. If we limit our attention to 99 per cent. of the population, even when n is comparatively small, we find regression on the mid-parent sensibly given by the Galtonian theory of the midparent, and when n is anything at all considerable, the Galtonian theory gives the value of the average offspring sensibly correct for 999 in 1000 of the population, becoming absolutely correct as n is indefinitely increased. The theory of regression shows us that the planes (ii) are the best fitting planes to the hyperboloids (i), and this fit becomes better and better as n becomes larger.

Except for very small values of n , it would probably be impossible to get an accurate test of the truth of the theory of the pure gamete by using formula (i) ; the errors of random sampling are for the great bulk of the population as large as the divergence between (i) and (ii).

(j). If a correlation table for parents and offspring be formed in a Mendelian population, while the regression will be real and linear, the variability of the number of allogenic couplets in the array of offspring due to a parent of definite allogenic constitution will steadily increase if we pass from the pure protogenic to the pure allogenic parent. The mean variability of the arrays of offspring is equal to $\sigma \sqrt{1 - r^2}$, where σ is the variability of the general population, and r the coefficient of parental correlation. This mean value exactly agrees with that given by statistical theory. But no one has hitherto observed this gradual change of variability in the arrays as we pass across a parental correlation table. It is not very marked, especially when n is large, and may, perhaps, have escaped notice ; still one would be rather surprised if it had. This change of variability of arrays of offspring seems to provide a method of finding n the number of couplets in the constitution of the zygote. This point will doubtless receive attention, and there is ample material already collected to test it upon.

(k). The frequency of an array of offspring due to a given parent depends upon the product of two skew binomials. Whether there is approach to any skew, or normal curve, when n is increased has not yet been investigated, but the deviation from normality exhibited would on the surface appear to be considerable, and such deviation would be inconsistent with the approximately elliptic contour lines which have been noted by Galton in discussing human characters.

(5). To sum up: The memoir shows that any physiological hypothesis, such as a generalised theory of the pure gamete, is not *à priori* inconsistent with the broad conceptions of linear regression, parental correlation, ancestral influence, and with the distribution of frequency actually observed in populations. The theory under discussion, indeed, leads to these results, and they have only escaped the Mendelians because they did not develop their formulæ for the case of a population crossing at random, but confined their attention too much to the hybridisation of pure races, and to self-fertilising individuals. On the other hand, there are marked numerical divergences between the results observed for populations and those deduced from the generalised theory of the pure gamete, which seem to definitely prevent its acceptance as a general theory of heredity. It would be of very great value to have a physiological theory which would not only lead up to the general laws of inheritance—involved in the principles of regression and of ancestral influence—this is actually achieved by the present theory—but which would also lead to numerical results in reasonable accord with experience. Such a theory will, no doubt, be discovered in time, but the touchstone of its truth will be that it gives, not only a formula which does not vary from one simple case of crossing to a second, but also general principles and numerical values in accordance with those which have been observed for large populations, whether they breed at random, or exhibit homogenic tendencies. I think we are justified in asking any future propounder of Mendelian formulæ to apply this touchstone, and to remember that the statistical view of inheritance is not at basis a theory, but a description of observed facts, with which any physiological theory must be in accord.
